AD-A114 534 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
ITERATIVELY REWEIGHTED LEAST SQUARES - ENCYCLOPEDIA ENTRY.(U)
FEB 82 D B RUBIN DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2328 NL

1 OF 1
AD
A114534

END
DATE
FILMED
6-82
DTIC

2

AD A114534

MRC Technical Summary Report # 2328

ITERATIVELY REWEIGHTED LEAST SQUARES - ENCYCLOPEDIA ENTRY

Donald B. Rubin

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1982

(Received October 27, 1981)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ITERATIVELY REWEIGHTED LEAST SQUARES - ENCYCLOPEDIA ENTRY

Donald B. Rubin

Technical Summary Report #2328

February 1982

ABSTRACT

Iteratively Reweighted Least Squares (IRLS) is a computationally attractive method for providing estimated regression coefficients that are relatively unaffected by extreme observations. Definitions and statistical justifications are reviewed, and a numerical example and a multivariate extension are included. This article is to be an entry in The Encyclopedia of Statistical Sciences.

AMS (MOS) Subject Classification: 6204, 62F35, 62H12, 62J05, 65F10

Key Words: Robust Regression, IRLS, EM Algorithm, Maximum Likelihood Estimation, M-Estimates, Weighted Regression, t-distribution

Work Unit Number 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

ITERATIVELY REWEIGHTED LEAST SQUARES - ENCYCLOPEDIA ENTRY

Donald B. Rubin

Iteratively reweighted least squares (IRLS) refers to an iterative procedure for estimating regression coefficients: at each iteration, weighted least squares computations are performed, where the weights change from iteration to iteration. Although IRLS has been used to estimate coefficients in nonlinear and logistic regressions, currently, IRLS tends to be associated with robust regression.

1. IRLS for Robust Regression

When using IRLS for robust regression, the weights are functions of the residuals from the previous iteration such that points with larger residuals receive relatively less weight than points with smaller residuals. Consequently, unusual points tend to receive less weight than typical points.

IRLS is a popular technique for obtaining estimated regression coefficients that are relatively unaffected by extreme observations. One reason for the popularity of IRLS is that it can be easily implemented using readily available least squares algorithms. Another reason is that it can be motivated from sound statistical principles (c.f. [1], [4], [9]). A third reason for its popularity is that some experience suggests it is a useful practical tool when applied to real data (c.f. [2], [7]). In order to define precisely IRLS for robust regression, some notation is needed.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

2. Weighted Least Squares Computations

Let Y be an $n \times 1$ data matrix of n observations of a dependent variable, let X be the associated $n \times p$ data matrix of n observations of p predictor variables, and let W be an $n \times n$ diagonal matrix of nonnegative weights, which for the moment we assume is fixed. Then the weighted least squares estimate of the regression coefficient of Y on X is given, as a function of W, by

$$b(W) = (X^T W X)^{-1} (X^T W Y) \quad , \tag{1}$$

if $(X^T W X)$ has rank p and is not defined otherwise.

Theoretical justification for the estimator $b(W)$ is straightforward. Suppose that for fixed W, the conditional distribution of Y given X has mean $X\beta$, where β is the $p \times 1$ regression coefficient to be estimated, and variance $\sigma^2 W^{-1}$, where σ^2 is the residual variance, usually also to be estimated. By noting that, for fixed W, $W^{1/2} Y$ has mean $W^{1/2} X\beta$ and variance $\sigma^2 I$, the standard Gauss-Markov arguments imply that $b(W)$ is the value of β that minimizes the residual sum of squares $(Y - X\beta)^T W(Y - X\beta)$ as well as the minimum variance unbiased estimator of β. If the conditional distribution of Y given X is normal for fixed W, then $b(W)$ is also the maximum likelihood estimate of β, and the associated maximum likelihood estimate of σ^2 is the weighted sum of squared residuals:

$$s(W)^2 = [Y - X\, b(W)]^T W [Y - X\, b(W)]/n \quad . \tag{2}$$

IRLS is used when the weight matrix is not fixed. Specifically, IRLS applies equation (1) to obtain $b^{(\ell+1)}$, the $(\ell+1)$st iterate of the regression coefficient, from the weight matrix of the previous iteration:

$$b^{(\ell+1)} = b(W^{(\ell)}) \quad . \tag{3}$$

In order to define a specific version of IRLS, we thus need only to define the weight matrix $W^{(\ell)}$.

3. The Weight Matrix and Iterations for Robust Regression

For Robust regression, the i^{th} diagonal element in the weight matrix $W^{(\ell)}$, $W_{ii}^{(\ell)}$, is a function $w(\cdot)$, of the i^{th} standardized residual obtained by using $b^{(\ell)}$ to predict Y_i:

$$W_{ii}^{(\ell)} = w(z_i) = w(-z_i) \tag{4}$$

where

$$z_i = (Y_i - X_i b^{(\ell)})/s^{(\ell)} \tag{5}$$

and $s^{(\ell)}$ is the estimate of σ at the ℓ^{th} iteration. A natural form for $s^{(\ell)}$ based on likelihood criteria is given by equation (2) with $W^{(\ell-1)}$ substituted for W, and thus, by equation (3), with $b^{(\ell)}$ substituted for $b(W)$:

$$s^{(\ell)} = s(W^{(\ell-1)}) \quad . \tag{6}$$

The scalar function $w(\cdot)$ in (4) is a nonnegative and nonincreasing monotone function and thus gives relatively smaller weight to points with larger residuals, e.g., $w(z) = 2/(1 + z^2)$.

With a specified form for $s^{(\ell)}$ and a specified form for the function $w(\cdot)$, IRLS proceeds by choosing a starting value $W^{(0)}$ e.g., the identity matrix, and then calculating $b^{(1)}$ from equations (1) and (3), $s^{(1)}$ from equations such as (2) and (6), and thence $W^{(1)}$ from equations (4) and (5); from $W^{(1)}$, the next iterates $b^{(2)}$, $s^{(2)}$ and $W^{(2)}$ are calculated; the procedure continues indefinitely unless some $s^{(\ell)} = 0$ or $X^T W^{(\ell)} X$ has rank less than p. Experience suggests that for many choices of weight functions, the iterations reliably converge.

4. Statistical Justifications for IRLS

A general statistical justification for IRLS for robust regression arises from the fact that it can be viewed as a process of successive substitution applied to the equations for M-estimates ([1],[2],[8],[9], [10]). Numerical behavior of IRLS for robust regression is considered in [3], [6], [10], [11].

A more specialized justification for IRLS, which is consistant with statistical principles of efficient estimation, arises from the fact that some M-estimates are maximum likelihood estimates under special distributional forms for the conditional distribution of Y given x. When M-estimates are maximum likelihood estimates, the associated IRLS algorithm is an EM-algorithm ([4], especially pp. 19-20), and consequently, general convergence results about EM algorithms apply to IRLS algorithms; important results are that each step of IRLS increases the likelihood and, under weak conditions, IRLS converges to a local maximum of the likelihood function. Details of the relationship between IRLS and EM, including general results on large and small sample rates of convergence, are given in [5].

5. IRLS/EM for the t-distribution

A specific example when IRLS is EM occurs when the specification for the conditional distribution of Y_i given X_i is a scaled t-distribution on r degrees of freedom. Then the associated weight function for IRLS is $w(z) = (r + 1)/(r + z^2)$, and the large sample rate of convergence for IRLS is $3/(r + 3)$. More generally, if $d(z)$ is the probability density function specified for the conditional distribution of Y_i given X_i, then the associated weight function is defined by

$$
\begin{aligned}
w(z) &= -d'(z)/zd(z) \quad \text{for} \quad z \neq 0 \\
&= \lim_{z \to 0} -d'(z)/zd(z) \quad \text{for} \quad z = 0 \ .
\end{aligned}
$$

A small numerical example is given in [5] and summarized here. Ten observations were drawn from a t-distribution on 3 degrees of freedom (-0.141, 0.678, -0.036, -0.350, -5.005, 0.886, 0.485, -4.154, 1.415, 1.546). The results of twenty steps of IRLS starting from $W^{(0)} = I$ are given in Table 1. The empirical rate of convergence for both $b^{(\ell)}$ and $s^{(\ell)}$ at the 20^{th} iteration is 0.6805 which agrees well with the theoretical small sample rate of convergence of 0.6806 as calculated in [5]; the large sample rate of convergence is 0.5. Since the rate of convergence of an EM algorithm is proportional to the fraction of information in the observed data (i.e., in Y and X in the robust regression context) relative to the information in the observed and missing data (i.e., in Y, X and W), we see that in this example the observed data have relatively more information about β and σ than is typical for samples of size ten from a t on three degrees of freedom. Further discussion of these points is given in [5].

Table 1.

Successive iterations of IRLS for example

Iteration-ℓ	$\beta^{(\ell)}$	$\sigma^{(\ell)2}$
1	-0.467496	1.537750
2	0.103069	1.673303
3	0.240781	1.603189
4	0.277822	1.524210
5	0.292411	1.466860
6	0.300280	1.427958
7	0.305188	1.401828
8	0.308413	1.384252
9	0.310571	1.372393
10	0.312027	1.364371
11	0.313012	1.358934
12	0.313680	1.355244
13	0.314133	1.352738
14	0.314442	1.351035
15	0.314651	1.349876
16	0.314794	1.349088
17	0.314890	1.348552
18	0.314956	1.348188
19	0.315001	1.347939
20	0.315032	1.347771

6. A Multivariate Extension

A potentially quite useful and simple generalization of the use of IRLS/EM for the t-distribution has apparently not yet appeared in the literature and illustrates the flexibility of IRLS. Suppose Y_i is q-variate and X_i is p-variate as before, where β is now $p \times q$, and let the conditional distribution of $Y_i - \beta X_i$ given X_i be a zero-centered linear transformation of a q-variate spherically symmetric t-distribution on r

degrees of freedom. Then the previous notation and equations apply with the following simple modifications: $b(W)$ defined by (1) is now $p \times q$, $s(W)^2$ defined by (2) is now $q \times q$, the weight function is given by

$$w(z_i) = (r + q)/(r + z_i^2) \tag{7}$$

where at the ℓ^{th} iteration

$$z_i^2 = (Y_i - X_i\, b^{(\ell)})[s^{(\ell)2}]^{-1}(Y_i - X_i\, b^{(\ell)})^T \quad . \tag{8}$$

IRLS begins with a starting value, $W^{(0)}$, e.g. the identity matrix, calculates the $p \times q$ matrix $b^{(1)}$ from equations (1) and (3), the $q \times q$ matrix $s^{(1)2}$ from equations (2) and (8), and thence the $n \times n$ diagonal matrix $W^{(1)}$ from equations (4), (7) and (8); $W^{(1)}$ leads to the next iterates $b^{(2)}$, $s^{(2)2}$, and so forth.

Under the t-specification, IRLS is EM and so each iteration increases the likelihood of the $p \times q$ location parameter β and the $q \times q$ scale parameter σ^2, and under weak conditions, the iterations will converge to maximum likelihood estimates of β and σ^2. IRLS thus provides a positive semi-definite estimate of the matrix of partial correlations among the q components of Y_i assuming the conditional distribution of Y_i given X_i is elliptically symmetric and long tailed (if r is chosen to be small). Some limited experience with real data suggests that this use of IRLS does yield estimates of correlation matrices rather unaffected by extreme observations.

REFERENCES

[1] Andrews, D. G., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). Robust Estimates of Location: Survey and Advances. Princeton University Press.

[2] Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series. Technometrics 16, 147-185.

[3] Byrd, R. H. and Pyne, D. A. (1979). ASA Procedings on Statistical Computations, 68-71.

[4] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). J. Roy. Statist. Soc. Ser. B#39, 1-38.

[5] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1980). Multivariate Analysis - V. North-Holland, 35-57.

[6] Dutter, R. (1977). J. Statist. Comput. Simul. 5, 207-238.

[7] Eddy, W. and Kadane, J. (1982). J. Amer. Statist. Assn.

[8] Holland, P. W. and Welsch, R. E. (1977). Commun. Statist. A#6, 813-827.

[9] Huber, P. J. (1964). Ann. Math. Statist. 35, 73-101.

[10] Huber, P. J. (1981). Robust Statistics. Wiley.

[11] Klein, R. and Yohai, V. J. (1981). Commun. Statist. A10, 2373-2388.

DBR/jvs

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER #2328	2. GOVT ACCESSION NO. AD-A114 534	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Iteratively Reweighted Least Squares - Encyclopedia Entry		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald B. Rubin		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709		12. REPORT DATE February 1982
		13. NUMBER OF PAGES 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust Regression, IRLS, Em Algorithm, Maximum Likelihood Estimation, M-Estimates, Weighted Regression, t-distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Iteratively Reweighted Least Squares (IRLS) is a computationally attractive method for providing estimated regression coefficient that are relatively unaffected by extreme observations. Definitions and statistical justifications are reviewed, and a numerical example and a multivariate extension are included. This article is to be an entry in The Encyclopedia of Statistical Sciences.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

ME

8